



City Research Online

City, University of London Institutional Repository

Citation: Baker, S.A. ORCID: 0000-0002-4921-2456 (2020). Tackling Misinformation and Disinformation in the Context of COVID-19. Cabinet Office C19 Seminar Series.,

This is the published version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/24612/>

Link to published version:

Copyright and reuse: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

City Research Online: <http://openaccess.city.ac.uk/> publications@city.ac.uk

Tackling Misinformation and Disinformation in the Context of COVID-19

Cabinet Office C19 Seminar Series

Dr Stephanie Alice Baker – City, University of London

One of the most significant challenges during the COVID-19 pandemic has been how to tackle the spread of misinformation and disinformation about the virus. Online news sources are interspersed with a series of false and misleading claims about the origin, transmission and alleged treatments for the virus. These range from innocuous claims to fraudulent products, and politically motivated attacks on scientists, governments and public authorities. Some of the misinformation shared online is genuinely believed to be true, while disinformation is intended to deceive and cause harm. Misinformation and disinformation present a serious risk to society by undermining public health messaging. The issue is not merely the volume of false and misleading information disseminated online, but that the overabundance of information about the virus has made it difficult to know who to trust and what information to believe. There are two major challenges raised by this 'infodemic', as the World Health Organisation (WHO) terms it: how does misinformation spread and how should governments and technology companies respond to it? Policy responses involving the latter require knowledge about the former.

How does misinformation spread?

Misinformation and disinformation flourish during times of crisis and uncertainty. Whether created intentionally or unintentionally, health misinformation thrives when people are concerned about their wellbeing and desperate for a miracle cure. Misinformation has existed for centuries, however prior to the internet false claims had limited reach. Digital technologies enable misinformation to spread at an unprecedented speed and scale, with social media allowing geographically distant communities to connect around common interests. Those who wish to spread misinformation now have the capacity to connect with networks of like-minded followers, who reinforce their beliefs. Together these networks intervene in public discourse and contribute to collective narratives about illness and disease.

Researchers have been examining the main claims and sources of misinformation related to COVID-19. In one study, I examined the viral spread of the conspiracy theory film, *Plandemic*. The 26-minute film promotes numerous falsehoods about COVID-19, many of which contradict authoritative advice from government healthcare agencies. Released on social

media on 4 May 2020, within days the film went viral acquiring over 8 million views. I analysed a sample of 30,815 tweets aggregated under the #Plandemic hashtag. The sample was collected from 7 May, when the #Plandemic hashtag was created, until 30 June 2020. The key findings include:

- Content creation: 30,815 tweets were created by 12,978 unique Twitter accounts. Most content was created by ordinary users with <1000 followers. Many of these accounts had a political agenda and affiliations to pro-Trump, conspiracy groups with the terms WWG1WGA, MAGA, KAG and QAnon listed in their bios. Although geo data indicated significant US representation among those using the hashtag, frequent sources of content creation occurred in the UK. 27% of the accounts that tweeted most frequently (>100 tweets) were created within one month of the film's release indicating strategic intent to promote the film and spread misinformation.
- Claims: The key claims disseminated using the hashtag fell into 4 main categories: 1) Deep State conspiracy theories that the pandemic is a hoax designed to control the population; 2) Politically motivated attacks directed towards public authorities and liberal politicians; 3) False claims about Bill Gates and the danger of vaccines; and 4) Critiques debunking the film. These claims were communicated in textual and visual form with questions, anecdotes, memes, images and videos used to engender suspicion and foster doubt.
- Amplification: 56% of the tweets with the greatest engagement (>100 retweets) were posted by influencers with over 30,000 followers. 25% of these were amplified by three accounts and 19% by verified accounts. Only 3% were posted by users <1000 followers.

These findings demonstrate that while ordinary users account for the majority of content creation, these accounts feed misinformation to influencers with large online followings. Despite QAnon conspiracy theorists dominating the hashtag, there were examples of misinformation moving beyond the margins with political candidates, lifestyle influencers and celebrities using their influence to amplify misinformation about the virus.

My research on the spread of 5G misinformation during the pandemic showed similar amplification by anti-globalist, conspiracy groups. However, verified accounts belonging to scientists and journalists intervened more frequently to debunk these narratives and relative to the #Plandemic hashtag, those who shared 5G conspiracy theories on Twitter were

geographically dispersed; geo data correlating with physical locations where 5G protests took place, highlighting the relationship between offline and online interactions in shaping the spread of misinformation. False and misleading claims regarding 5G's alleged association with COVID-19 seemed to engage a more general public audience, compared to the seemingly co-ordinated efforts to amplify content using the #Plandemic hashtag. The different motivations used to spread these false and misleading claims, emphasises the importance of employing different interventions to tackle misinformation and disinformation.

How to tackle misinformation and disinformation?

Technology companies have responded to these concerns by elevating authoritative content on their platforms and sharing critical updates from government healthcare agencies when users log onto their sites and conduct searches related to COVID-19. This co-ordinated effort has also resulted in many social media companies updating their rules and policies around the concept of harm. While content moderation is relatively straightforward when it comes to responding to risks of imminent physical harm (e.g. removing false claims advocating harmful practices such as drinking bleach or colloidal silver as a treatment for COVID-19), using the concept of harm to tackle misinformation about an emerging pandemic is more complicated.ⁱ Not only does much remain unknown about the virus, public health advice is highly politicised and subject to change as exemplified by the WHO's statements about the transmission of COVID-19 and the wearing of face masks as a precautionary device. There are several important policy questions to consider:

1. How should we respond to the spread of what appears to be false or misleading information during a pandemic when much remains uncertain about the virus? When there is a lack of scientific consensus, whose advice should we elevate?
2. How might we contend with misinformation when presented as questions, anecdotal evidence or in visual form? Should these communicative techniques be treated in the same way as overt claims, and evaluated on the basis of their perceived intent or potential consequences?
3. In light of these considerations, what are the most effective strategies to moderate misinformation and disinformation online? e.g. removal, demotion, labels, flagging. This question is complicated by the propensity for community flagging and search queries to be gamed and the difficulty of enacting these strategies in real-time and at scale.

These considerations problematise using binary conceptions of truth and falsity as the basis for content moderation policies.

Tackling health misinformation effectively requires responsibility from technology companies, news organisations and users to limit the spread of false and misleading advice, but it also requires government oversight. Despite their coordinated efforts, the content moderation strategies adopted by technology companies in tackling misinformation during the pandemic vary considerably. My research showed a significant discrepancy between the content moderation policies promised and practiced by the major social media companies. Strategies need to be implemented for those forms of misinformation that evade fact checking (e.g. memes, questions, anecdotes) with research on health misinformation demonstrating the power of these communicative techniques to inform what information people trust and who they believe.ⁱⁱ For authoritative advice to be trusted, governments and politicians need to be transparent about what is known and what is unknown given that unclear and contradictory guidelines are likely to erode trust in government healthcare agencies. Finally, while elevating authoritative content may prove useful for those inclined to trust public authorities, it is unlikely to be effective for those with a deep distrust of experts and political elites. Policy responses designed to limit the spread of misinformation must acknowledge the role of trust in shaping what information people choose to believe.

Dr Stephanie Alice Baker is a Senior Lecturer in Sociology at City, University of London. Her research explores how communities connect and communicate online, particularly around issues pertaining to health.

Cite this article as: Baker, S.A. (2020). Tackling Misinformation and Disinformation in the Context of COVID-19. *Cabinet Office C19 Seminar Series*.

ⁱ Baker, S.A., Wade, M. & Walsh M.J. (2020). The Challenges of Responding to Misinformation during a Pandemic: content moderation and the limitations of the concept of harm. *Media International Australia – Extraordinary Issue: Coronavirus, Crisis and Communication*.

ⁱⁱ Baker, S.A. & Rojek, C. (2020). *Lifestyle Gurus: Constructing Authority and Influence Online*. Polity: Cambridge.